

Open Science for Life in Space: Data Sharing and Tools for Knowledge Discovery

Ryan T. Scott¹, Lovorka Degoricija¹, Lauren M. Sanders², Danielle K. Lopez¹, Samrawit G. Gebre¹, Ahleah Rohr Daniel³, and Sylvain V. Costes³

¹KBR, Moffett Field, CA; ²Blue Marble Space Institute of Science, Moffett Field, CA; ³Space Biosciences Division, NASA Ames Research Center, Moffett Field, CA

The next era in human space exploration is rapidly approaching and will require the use of countermeasures to deep space health hazards. The development of countermeasures (or, the re-purposing of existing agents) will be highly dependent on our understanding of basic biological responses to space stressors (e.g. ionizing radiation, altered gravitational fields, altered day-night cycles, confinement, isolation, hostile-closed environments, distance-duration from Earth, exposure to celestial regolith, etc.). The fast-growing array of space biological data, which in the past was simply archived after minimal analysis, holds great potential if it can be reorganized and formatted for Open Science. Organizing the data for such analysis is a challenge because of its diverse nature (molecular, cellular, tissue, imaging, whole organism and behavior). We will discuss here several strategies that NASA's Biological and Physical Science Division has put in place to maximize the return on investment for spaceflight bioscience data.

Open Science, as a scientific philosophy, is the concept that the more people who have access to the data, the more knowledge will be gained from it. This guiding principle led NASA to develop GeneLab in 2015. GeneLab houses spaceflight and relevant ground-based multi-omics data, and has grown to ~400 transcriptomic, proteomic, metabolomic and epigenomic datasets from plant, rodent, small animal, and microbial space experiments. GeneLab provides users with various tools for data analysis and a visualization portal that allows users to interact with gene expression data from space-related 'omics experiments. Open Science is also about building scientific communities, and with this spirit in mind, GeneLab has spawned several Analysis Working Groups (AWGs), comprised of more than 200 volunteer scientists. The AWGs initially provided feedback on the processing pipeline and metadata 'omics standards for GeneLab. Over the last few years, they have become a community-driven science enterprise, engaging in large meta-analysis of GeneLab datasets, resulting in 10 publications (beyond the originally submitted research). Overall, the Open Science nature of GeneLab has resulted in a high degree of data re-use, resulting in 38 additional publications derived from the original 67 publication over the past four years.

The enormous success and knowledge gained from GeneLab has led to a collection of sister NASA "Open Science Data Repositories (OSDR)" and research support groups. These include the NASA Ames Life Sciences Data Archive (ALSDA), the NASA Biological Institutional Scientific Collection (NBISC), and the Biospecimen Sharing Program (BSP). All are adopting the GeneLab data architecture system to maximize open-access, find-ability, accessibility, interoperability, and reusability (FAIR). ALSDA collects and curates phenotypic-physiological bioimaging-behavioral data from space and space-relevant non-human experiments, oftentimes coming from the same omics-associated experimental datasets found in GeneLab. Since 2021, a community of ~100 researchers have rallied around ALSDA, to provide feedback in a new ALSDA AWG focused on phenotypic-physiological investigation-sample-assay metadata standards (e.g., Micro-Computed Tomography, Light/Fluorescence Microscopy, Western Blot, Flow Cytometry, Novel Object Recognition, Elevated Plus Maze, etc. of ~50 assays collected). These standards are part of a new single point-of-entry data submission portal for all non-human Space Biology and Human Research Program principal investigators, to submit, curate, and share their research data.

With open-access space biological data now collected and curated together with rich metadata, and with the potential for linkage to "big data" from the international biological and medical communities (NIH, EBI, etc.), the artificial intelligence and machine learning (AI/ML) era has started for Space Biology. Several other talks will cover these topics in this conference.